M&G 10552.26-US-01

COMPOSITION AND METHOD FOR DETECTING MUTAGENS

Field of the Invention

The present invention relates to methods and compositions for detecting a mutagen. The compositions include a DNA construct, an expression vector, and a host cell including a mutagen sensitive gene operably linked to a fluorescent protein. The method includes exposing a host cell including a mutagen sensitive gene operably linked to a fluorescent protein and monitoring the fluorescent protein.

Background of the Invention

The current increased awareness of environmental contamination by diverse classes of chemicals has led to increasing concern about the impact of this contamination on human health. Some of these environmental contaminants may be toxic and/or carcinogenic, which emphasizes the need for rapid and inexpensive screening methods for the hazards of environmental contaminants. Many existing tests for carcinogens are based on animal models or tissue culture techniques, which are both time-consuming and costly. These animal and tissue culture tests can be supplemented with simpler tests on bacterial cell cultures. However, the most widespread bacterial mutagenicity the test, the Ames test, requires repeated culturing of many cell samples, lengthy incubation, and tedious analysis. These consume time, money, and reagents. There exists a need for simpler and more robust bacterial or cell culture tests for mutagens.

Summary of the Invention

The present invention relates to methods and compositions for detecting a mutagen. The compositions include a DNA construct, an expression vector, and a host cell each including a mutagen sensitive gene operably linked to a fluorescent protein. The method includes exposing a host cell including a mutagen sensitive gene operably linked to a fluorescent protein and monitoring the fluorescent protein.

The present invention includes a DNA construct including a mutagen sensitive gene operably linked to the coding sequence for a fluorescent protein. In a preferred embodiment, the mutagen sensitive gene includes an SOS gene. The SOS gene can include one or more promoters or coding sequences whose expression is induced or up regulated in a cellular response to a mutagen. A preferred SOS gene is derived from *E. coli* or *S. typhimurium* and includes a *umuC* gene, a *umuD* gene, a control sequence for either of these genes, or a combination thereof. The DNA construct of the invention can include the coding sequence for any of a variety of fluorescent proteins. Preferably, the fluorescent protein is a green fluorescent protein, preferably from the jellyfish *Aequorea victoria*. The coding sequence for the fluorescent protein can encode a variant green fluorescent protein. The mutagen sensitive gene and/or the fluorescent protein can be a naturally occurring or variant gene or protein. The expression construct can encode a protein of the invention, a polypeptide including an amino acid sequence of a UmuD protein, a UmuC protein, or a combination thereof and an amino acid sequence of a fluorescent protein.

The invention also includes an expression vector including a mutagen sensitive gene operably linked to a coding sequence for a fluorescent protein. The expression vector can include the DNA construct of the invention. Preferably, the expression vector includes an SOS gene, or variant thereof, derived from plasmid pSE 117 and a coding sequence for a fluorescent protein. A preferred plasmid expression vector is pTJgfp.

In another embodiment, the invention includes a host cell including a mutagen sensitive gene operably linked to a coding sequence for a fluorescent protein. Preferred host cells include *S. typhimurium* and *E. coli*. The host cell can include a DNA construct or expression vector of the invention. Alternatively, the host cell can have its genomic DNA altered to include a heterologous mutagen sensitive gene operably linked to a homologous coding sequence for a fluorescent protein or to include a heterologous coding sequence for a fluorescent protein operably linked to a homologous mutation sensitive gene.

In one embodiment, the invention includes a method of determining or detecting a mutagen. The method includes contacting a test compound with a host cell of the invention, monitoring the host cell for the fluorescent protein, and, when an amount or distribution of the fluorescent protein meets or exceeds a predetermined threshold value, determining that

the test compound is a mutagen. The method can employ a host cell either at stationary phase or at logarithmic phase. Contact with the test compound can deplete a nutrient, or even starve the host cell, and the method will still robustly detect a mutagen. Preferably, when contacting depletes a nutrient or starves a cell, the cells are in a logarithmic phase.

Monitoring fluorescent protein typically includes detecting fluorescence, such as with a fluorescence detector that reads a microtiter plate. The mutagen can be detected through increases in the level or output of fluorescence compared to a control sample. Alternatively, the mutagen can be determined or detected by statistically analyzing the amount of green fluorescent protein in replicate wells. The statistical analysis can include analysis of the location, shape, or distribution of fluorescent output, cells, or other data.

Brief Description of the Figures

Figure 1 illustrates construction of a *umuC'::gfp* fusion plasmid. A *Hind*III-*Eco*RI fragment from pSE117 was ligated to a 700bp *gfp* gene which was obtained via PCR amplification using *Hind*III and *Eco*RI primers.

Figures 2a and 2b illustrate fluorescence of *E. coli* GW2100/pTJ*gfp* versus time following treatment with *N*-methyl-*N*-nitro-*N*-nitrosoguanidine (MNNG): (a) stationary phase cells; (b) logarithmic phase cells;□, 0 μg/ml MNNG; Δ, 0.1 μg/ml MNNG; X, 1.5 μg/ml MNNG; ♦. 3.5 μg/ml MNNG; O, 7 μg/ml MNNG.

Figures 3a and 3b illustrate fluorescence of *E. coli* GW2100/pTJ*gfp* versus time following treatment with methylmethane sulphonate (MMS): (a) stationary phase cells; (b) logarithmic phase cells; □, 0 μg/ml MMS; Δ, 13 μg/ml; X, 325 μg/ml; O, 650 μg/ml; ♦, 1300 μg/ml.

Figures 4a and 4b illustrate fluorescence of *E. coli* GW2100/pTJ*gfp* versus time after 254 nm UV irradiation: (a) stationary phase cells; (b) logarithmic phase cells; □, 0 J/m^2 UV; Δ, 1 J/m^2 UV; X, 3 J/m^2 UV; O, 6 J/m^2 UV.

Figure 5 illustrates fluorescence of *E. coli* GW2100/pTJ*gfp* cells in logarithmic phase versus time following treatment with MNNG in 0.85% saline: □, 0 μg/ml MNNG; Δ, 0.1 μg/ml MNNG; X, 1.5 μg/ml MNNG; O, 3.5 μg/ml MNNG; ♦, 7 μg/ml MNNG.

Figure 6 illustrates fluorescence of *E. coli* GW2100/pTJ*gfp* cells in stationary phase versus time following treatment with MNNG in 0.85% saline: □, 0 μg/ml MNNG; Δ, 0.1 μg/ml MNNG; X, 1.5 μg/ml MNNG; ♦, 3.5 μg/ml MNNG; O, 7 μg/ml MNNG.

Figure 7 illustrates a histogram representing mutant colony numbers from 40 parallel cultures following exposure to MNNG: 0 μg/ml, 0.1μg/ml, 3.5μg/ml.

Figure 8 illustrates a histogram representing mutant colony numbers from 37 parallel cultures following exposure to MMS: 0 μg/ml, 13μg/ml, 325μg/ml.

Figures 9a and 9b illustrate histograms representing mutant colony numbers and fluorescence emission from 24 parallel cultures following exposure to MNNG: 0 μg/ml, 1.7μg/ml; a.) fluorescence emission values; b.) revertant colony numbers.

Figure 10 illustrates a histogram representing fluorescence emission from 84 parallel cultures following exposure to MMS: 0 μg/ml, 13μg/ml, 325μg/ml.

Figure 11 illustrates a histogram representing fluorescence emission from 84 parallel cultures following exposure to MNNG: 0 μg/ml, 0.1μg/ml, 3.5μg/ml.

Figure 12 illustrates a histogram representing fluorescence emission from 84 parallel cultures following exposure to 254nm UV irradiation: 0 J/m^2, 1 J/m^2, 3 J/m^2.

Figure 13 illustrates a nucleotide sequence of a preferred SOS gene of the invention.

Figure 14 illustrates a coding sequence for and the amino acid sequence of a preferred fluorescent protein of the invention.

Figure 15 illustrates a preferred DNA construct of the invention.

Detailed Description of the Invention

Definitions

As used herein, "test compound" refers to a compound to be tested for mutagenicity. The test compound can be a single compound, a mixture of compounds, even a mixture of biological materials, such as a plant extract, soil, fresh water, salt water, or the like. The test compound can also be a mixture of test compounds. The test compound can be applied to,

mixed with, or contacted with a host cell of the invention in one or more of a variety of physical states. For example, the test compound may be a component of a solution that is mixed with the nutrient media for the host cell. Alternatively, the test compound may include an insoluble solid that can be suspended in or contacted with the nutrient medium. The test compound can even be a gas that is bubbled through the nutrient medium containing host cell.

As used herein, "mutagen" refers to a physical or chemical agent that is capable of increasing frequency of mutation above the spontaneous, background level. A mutagen can induce mutations in DNA and in living cells or cause biochemical damage to a gene. "Mutation" refers to the process by which genetic material undergoes a detectable and/or heritable structural change, or the result of such a change.

As used herein, "fluorescence", "fluoresce", or "fluorescent" refers to a type of luminescence in which an atom or molecule emits visible radiation in passing from a higher to a lower electronic state. These terms are restricted to phenomena in which the time interval between absorption and emission of energy is extremely short (10^{-8} to 10^{-3} second). This distinguishes fluorescence from phosphorescence, in which the time interval may extend to several hours. Fluorescence can result from absorbing energy derived from exciting radiation of a wavelength that is usually shorter than the wavelength of the light emitted. Fluorescence can also arise from striking an atom or molecule with a subatomic particle. As used herein, fluorescence is distinct from "bioluminescence" or "chemiluminescence". Bioluminescence refers to the production of light by certain enzyme catalyzed reactions in living organisms. Chemiluminescence refers to the production of visible light occurring as a result of a chemical reaction. A fluorescent protein need not undergo an enzyme catalyzed or chemical reaction to fluoresce.

As used herein, "expression vector" means a DNA construct including a DNA sequence (e.g., a sequence encoding a fluorescent protein) which is operably linked to a suitable control sequence (e.g. all or part of a mutagen sensitive gene) capable of affecting the expression of the DNA in a suitable host. Such control sequences may include a promoter to affect transcription, an optional operator sequence to control transcription, a sequence encoding suitable ribosome-binding sites on the mRNA, and sequences which

control termination of transcription and translation. Different cell types may be employed with different expression vectors. The vector may be a plasmid, a phage particle, or simply a potential genomic insert. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, under suitable conditions, integrate into the genome itself. In the present specification, plasmid and vector are sometimes used interchangeably. However, the invention is intended to include other forms of expression vectors which serve equivalent functions and which are, or become, known in the art. Useful expression vectors, for example, can include segments of chromosomal, non-chromosomal and synthetic DNA sequences such as various known derivatives of known bacterial plasmids, e.g., plasmids from *E. coli* including Col E1, pCR1, pBR322, pMb9, pUC 19 and their derivatives, wider host range plasmids, e.g., RP4, phage DNAs e.g., the numerous derivatives of phage λ, e.g., NM989, and other DNA phages, e.g., M13 and filamentous single stranded DNA phages, yeast plasmids such as the 2μ plasmid or derivatives thereof, vectors useful in eukaryotic cells, such as vectors useful in animal cells and vectors derived from combinations of plasmids and phage DNAs, such as plasmids which have been modified to employ phage DNA or other expression control sequences. Expression techniques using the expression vectors of the present invention are known in the art and are described generally in, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Press (1989).

As used herein, "signal sequence" means a sequence of amino acids bound to the N-terminal portion of a protein which facilitates the secretion of the mature form of the protein outside of the cell. This definition of a signal sequence is a functional one. The mature form of the extracellular protein lacks the signal sequence which is cleaved off during the secretion process.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with

another nucleic acid sequence. For example, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" fluorescent protein nucleic acid molecule or mutagen sensitive gene nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the nucleic acid. An isolated fluorescent protein nucleic acid molecule or mutagen sensitive gene nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated fluorescent protein nucleic acid molecules or mutagen sensitive gene nucleic acid molecules therefore are distinguished from the corresponding nucleic acid molecules as they exist in natural cells. However, an isolated fluorescent protein nucleic acid molecule or mutagen sensitive gene nucleic acid molecule includes fluorescent protein nucleic acid molecules or mutagen sensitive gene nucleic acid molecules contained in cells that ordinarily express fluorescent protein or mutagen sensitive gene where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

As used herein, "percent (%) sequence identity" with respect to the amino acid or nucleotides sequences identified herein is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in a fluorescent protein sequence or a mutagen sensitive gene sequence, after

aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity.

Methods for performing sequence alignment and determining sequence identity are known to the skilled artisan, may be performed without undue experimentation, and calculations of identity values may be obtained with definiteness. See, for example, Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 19 (Greene Publishing and Wiley-Interscience, New York); and the ALIGN program (Dayhoff (1978) in *Atlas of Protein Sequence and Structure* 5:Suppl. 3 (National Biomedical Research Foundation, Washington, D.C.). A number of algorithms are available for aligning sequences and determining sequence identity and include, for example, the homology alignment algorithm of Needleman *et al.* (1970) *J. Mol. Biol. 48*:443; the local homology algorithm of Smith *et al.* (1981) *Adv. Appl. Math. 2*:482; the search for similarity method of Pearson *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:2444; the Smith-Waterman algorithm (*Meth. Mol. Biol. 70*:173-187 (1997); and BLASTP, BLASTN, and BLASTX algorithms (see Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-410). Computerized programs using these algorithms are also available, and include, but are not limited to: ALIGN or Megalign (DNASTAR) software, or WU-BLAST-2 (Altschul et al., Meth. Enzym., 266:460-480 (1996)); or GAP, BESTFIT, BLAST Altschul et al., supra, FASTA, and TFASTA, available in the Genetics Computing Group (GCG) package, Version 8, Madison, Wisconsin, USA; and CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California. Those skilled in the art can determine appropriate parameters for measuring alignment, including algorithms needed to achieve maximal alignment over the length of the sequences being compared. Preferably, the sequence identity is determined using the default parameters determined by the program. Specifically, sequence identity can determined by the Smith-Waterman homology search algorithm (*Meth. Mol. Biol.* 70:173-187 (1997)) as implemented in MSPRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty of 12, and gap extension penalty of 1. Preferably, paired amino acid comparisons can be carried out using the GAP program of the GCG sequence analysis software package of Genetics Computer Group, Inc., Madison, Wisconsin, employing the

blosum62 amino acid substitution matrix, with a gap weight of 12 and a length weight of 2.

With respect to optimal alignment of two amino acid sequences, the contiguous segment of the variant amino acid sequence may have additional amino acid residues or deleted amino acid residues with respect to the reference amino acid sequence. The contiguous segment used for comparison to the reference amino acid sequence will include at least 20 contiguous amino acid residues, and may be 30, 40, 50, or more amino acid residues. Corrections for increased sequence identity associated with inclusion of gaps in the derivative's amino acid sequence can be made by assigning gap penalties.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature that can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel *et al.*, Current Protocols in Molecular Biology (Wiley Interscience Publishers, 1995).

"Stringent conditions" or "high-stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example, 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50 °C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42 °C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42 °C, with washes at 42 °C in 0.2 x SSC (sodium

chloride/sodium citrate) and 50% formamide at 55 °C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55 °C.

"Moderately-stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Press, 1989), and include the use of washing solution and hybridization conditions (*e.g.*, temperature, ionic strength, and %SDS) less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37 °C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/mL denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50 °C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

Mutagen Sensitive Gene

As used herein, "mutagen sensitive gene" refers to any gene that responds to the presence of or action of a mutagen by altering the expression of one or more gene products. For example, a mutagen sensitive gene can increase production of a protein in response to mutagen induced DNA damage. Mutagen sensitive genes include genes having nucleic acid sequences that are either naturally occurring (wild-type) or variants of these sequences. The mutagen sensitive gene of the invention can respond to a mutagen by any of a variety of mechanisms. Preferably, the mutagen sensitive gene induces production of a protein.

A mutagen sensitive gene can be isolated from nature or can be produced by recombinant or synthetic means. The term specifically encompasses naturally-occurring isoforms or alleles of the mutagen sensitive gene, naturally-occurring variant forms, and the like. Methods for detecting, isolating, and characterizing mutagen sensitive genes are well known in the art.

SOS Gene

SOS repair or error prone repair occurs in bacteria and helps them recover from potentially lethal stresses, such as exposure to a mutagen. The SOS system mediates various

cellular responses including mutagenesis, activated excision repair, and activation of latent phage genomes. Although not limiting to the present invention, mutagenesis can occur by filling gaps opposite thymine dimers by replication rather than by daughter-strand transfer. Such replication is typically inaccurate. Expression of many of the genes in the SOS regulon is controlled by the LexA protein. LexA acts as a transcriptional repressor of these unlinked genes by binding to specific sequences (LexA boxes) located within the promoter region of each LexA-regulated gene. Alignment of 20 LexA binding sites found in the *E. coli* chromosome reveals a consensus sequence of 5'-TACTG(TA)(5)CAGTA-3'. DNA sequences that exhibit a close match to the consensus sequence exhibit a low heterology index and bind LexA tightly, whereas those that are more diverged have a high heterology index and are not expected to bind LexA.

SOS-responsive genes include genes that are switched on or whose transcription is induced by the SOS repair system and control sequences for these genes. Genes switched on or induced in the SOS response or their control or coding sequences are referred to herein as "SOS genes". SOS genes of several bacteria, including *Escherichia coli* and *Salmonella typhimurium*, are known in the art and have been isolated and characterized. This is described in references including: Walker, G.C., The SOS response of *Escherichia coli*, Chapter 84 in *Escherichia coli* and *Salmonella typhimurium* : cellular and molecular biology (1987) ASM Press, Washington D.C., USA. Sanderson K.E. et al., Microbiological Reviews 59(2):241-303 (1995). Berlyn M.K.B., Linkage map of Escherichia coli K-12, edition 10: the traditional map, Microbiology & Molecular Biology Review (Washington, DC) 62(3):814-984 (1998).

SOS genes include a umu gene, such as a *umuC* gene, a *umuD* gene, or a *umuDC* operon, which can be found in *E. coli* and *S. typhimurium*, and a muc gene, such as a *mucA* gene, a *mucB* gene, or a *mucAB* operon, which can be found on transferable plasmids. SOS genes also include a *col* gene, such as *colE1*; a din gene, such as a *dinA*, a *dinB*, a *dinD*, a *dinD1::Mud*(Ap,lac), a *dinF*, a *dinG*, a *dinH*, a *dinI*, or a *dinY* gene; an *imp* gene, such as an *impA* gene, an *impB* gene, an *impC* gene, or an *impCAB* operon; a *lex* gene, such as a *lexA* gene; an *alk* gene, such as an *alkA'* gene; a *pri* gene, such as a *priA* gene; a *rec* gene, such as a *recA*, a *recA'*, a *recF*, a *recL152*, or a *recN* gene; a *rer* gene; a *ruv* gene, such as a *ruvA* or

ruvB gene; a *sam* gene, such as a *samA* gene, *a samB* gene or a *samAB* operon; a sfi gene, such as a *sfiA* gene; a *ssb* gene; a *sul* gene, such as a *sulA* gene; and the like. See, e.g., Walker, G.C., Microbiol. Rev. 48(1)60-93 (1984); Langer et al., J. Bacteriol. 145, 1310-1316 (1981); Elledge et al., J. Bacteriol. 155, 1307-1315 (1983); Perry et al., Nature 300, 278-281 (1982); and H. Shinagawa et al., Gene, 23, 167 (1983).

SOS genes include mutants of these genes, homologs of these genes, and genes that complement these genes. Such mutant, homologous, and complementary genes can be isolated from bacteria, plasmids, and bacteriophages. Suitable *umuC* mutants include those identified by Woodgate R. et al. Journal of Bacteriology 176(16):5011-21 (1994). Suitable complementary genes can be isolated from a plasmid that complements a chromosomal gene. Such complementary genes include mucAB-like gene sequences and umu-complementing operons from plasmids R391, R446b, or R471a.

SOS genes, mutants of these genes, homologs of these genes, complements of these genes and the like can be isolated from a variety of organisms, in particular *E. coli* and *S. typhimurium*, and also *Bacillus* such as *B. subtilis*, *Deinococcus* such as *D. radiodurans*, *Erwinia* such as *E. chrysanthemi*, *Lactococcus* such as *L. lactis*, *Neisseria*, *Paracoccus* such as *P. denitrificans*, *Pseudomonas* such as *P. aeruginosa*, *Rhodobacter* such as *R. sphaeroides*, *Staphylococcus* such as *S. epidermidis*, *Streptococcus* such as *S. pneumoniae* and *S. mutans*, *Streptomyces* such as *S. coelicolor*, and *Sulfolobus* such as *S. solfataricus*.

The following sequences can provide an SOS gene for the DNA construct of the invention. The nucleotide and corresponding amino acid sequences for an *E. coli umuDC* operon encoding proteins functional in UV mutagenesis and including a promoter for this operon have been reported by Perry et al., Proc. Natl. Acad. Sci. USA 82, 4331-4335 (1985) and given GenBank accession number M13387 (SEQ ID NO: A) and also by Kitigawa et al. *supra* (GenBank accession number M10107). The nucleotide and corresponding amino acid sequences for a *S. typhimurium* plasmid R46 encoding a *mucAB* gene for mucA and mucB proteins have been reported by Hall et al. and given GenBank accession number X16596 (SEQ ID NO: D). See also, Kulaeva O.I. et al. J. Bacteriol. 177(10):2737-2743 (1995). Plasmid R46 is the parent plasmid for plasmid pKM101. The nucleotide and corresponding amino acid sequences for the *E. coli* plasmid pKM101 encoding at least muc genes have been

reported by Perry, K.L. et al. and by Tanooka et al. (Proc. Nat. Acad. Sci. USA 82(13):4331-4335 (1985) and J. Bacteriol. 173(9):2906-2914 (1991), respectively) and given GenBank accession numbers D90147 (SEQ ID NO: C), M13388, and M12287. The nucleotide and corresponding amino acid sequences for a *S. typhimurium* plasmid R394 encoding *mucA* and *mucB* genes have been reported by Woodgate et al. and given GenBank accession number AF039836 (SEQ ID NO: B). The nucleotide and corresponding amino acid sequences for a *S. typhimurium* LT2 *umuDC* operon have been reported by Smith et al. and Thomas et al. (J. Bacteriol. 172:4694-4978 (1990) and J. Bacteriol. 172:4979-4989 (1990), respectively) and given GenBank accession numbers M57431 and M35010 (SEQ ID NO: G). See also, Nohmi T. et al., J. Bacteriol. 173(3):1051-63 (1991). The disclosures of each of the GenBank accessions mentioned in this paragraph and the Perry, K.L. et al. reference are incorporated herein by reference.

The following sequences of *imp* genes can also provide an SOS gene for the DNA construct of the invention. The sequence of an *impCAB* operon has been reported by Lodwick D. et al. (Nucleic Acids Research 18(17):5045-50 (1990)). Sequences of impCAB-like genes from plasmids have been reported by Runyen-Janecky LJ. et al. and Lodwick D. et al. (Infection & Immunity 67(3):1415-1423 (1999) and Molecular & General Genetics 229(1):27-30 (1991), respectively)

Preferred SOS genes include the *umuC* gene and the *umuD* gene with the corresponding control sequence, preferably of *Escherichia coli*. The nucleotide and corresponding amino acid sequences of these genes are shown in Figure 13 (SEQ ID NO: A) and have been reported by Perry et al., Proc. Natl. Acad. Sci. USA 82, 4331-4335 (1985) and given GenBank accession number M13387 (SEQ ID NO: A). The *umuC* gene and the *umuD* gene are expressed under control of their natural promoter. One promoter controls the expression of both genes. This promoter is located upstream of the *umuD* and *umuC* coding sequences and, although not limiting to the present invention, is negatively regulated by the LexA protein. After DNA damage, LexA protein can be cleaved and the gene activated. The UmuD and UmuC proteins form a novel polymerase that provides the cell with the capacity to polymerize opposite DNA damage lesions.

A fragment of this sequence including nucleotides 1 to 968 of SEQ ID NO: A can be

employed in the present constructs and methods as an SOS gene that responds to mutagens and powers expression of a fluorescent protein. A coding sequence for a heterologous protein, such as a fluorescent protein, can be expressed when inserted in place of all of part of the coding sequence of the *umuC* gene and/or the *umuD* gene, or inserted into and in reading frame with either of these coding sequences. Preferably, a heterologous coding sequence is placed into the *umuC* gene coding sequence at a location such as after nucleotide 968 of SEQ ID NO: A.

Additional SOS-Like Genes

The organisms listed above include other genes for inducible DNA repair responses, such as the adaptive response, which can also be employed in the constructs and methods of the present invention. In addition, homologues and similar-function genes to the SOS genes of bacteria that have been found in the lower eukaryotes. These include the *rad* (e.g. *rad2*, *rad27*, *rad30*, *rad54*, and *rad52*), *din* (e.g. *din7*), *ntg* (e.g. *ntg1* and *ntg2*), snm (e.g. snm1), rev (e.g. rev2), exonuclease-1, and small subunit of ribonucleotide reductase genes of the yeast *Saccharomyces cerevisiae*. These also include the *rad* and *rph16+* genes of the yeast *Schizosaccharomyces pombe*. See, e.g., Wolter R. et al. Molecular & General Genetics 250(2):162-8 (1996) and Mitchel R.E. et al. Mutation Research 183(2):149 (1987). The fission yeast include suitable genes in their UVDR DNA repair pathway. SOS-like genes can also be found in fungi including *Neurospora crassa, Ustilago maydis*, *Aspergillus sp* (e.g. *spergillus nidulans*), and protists. See, Lee M.G. et al. Molecular & General Genetics 185(2):245-50 (1982). Further, the eukaryote *Drosophila melanogaster* includes an inducible repair-recombination system in their female germ line.

Additional Mutagen Sensitive Genes

Suitable mutagen sensitive genes include a variety of genes whose activity is known to be inducible in the presence of or in response to mutagens. These include a *dam* gene; a *mut* gene, such as a *mutD*, a *mutH*, a *mutL*, or a *mutS* gene; a *dna* gene, such as a *dnaB* or a *dnaQ* gene; an *uvr* gene, such as an *uvrA'*, an *uvrC*, an *uvrD*, or an *uvrD252* gene; a gene for the sliding clamp subunit, or another subunit, of DNA polymerase III; a *hex* gene from

Streptococcus pneumoniae; and the like. These genes, or homologous genes, can be found in *E. coli*, *S. typhimurium*, and additional microbes listed above for SOS genes. Such genes have been characterized and their sequences are known in the art. See: Kleinsteuber S. et al., Molecular & General Genetics 248(6):695-702 (1995); Quinones A. et al., Molecular & General Genetics 211(1):106-12 (1988); Martinez A. et al. J. Bacteriol. 179(16):5188-94 (1987); Siegel E.C. Molecular & General Genetics 191(3):397-400 (1983); and Prudhomme M. et al., J. Bacteriol. 173(22):7196-203 (1991). Suitable genes also include those essential during inducible mutagenesis. Such genes include *polA* (DNA polymerase I) and the genes for DNA ligase, glycosylases, and apurinic/apyrimidinic endonucleases. The present constructs and methods can also employ homologs of these genes as found in another bacterium or bacteriophage species.

Variant Mutagen Sensitive Gene

In addition to the native sequence mutagen sensitive gene described herein, it is contemplated that variants of the mutagen sensitive gene can be prepared. A variant mutant sensitive gene retains its sensitivity or responsiveness to mutagens. A variant mutagen sensitive gene can be prepared by introducing appropriate nucleotide changes into the native gene, or by synthesis of the variant gene. A preferred variant mutagen sensitive gene has at least about 80% nucleic acid sequence identity with, preferably at least about 85% nucleic acid sequence identity with, more preferably at least about 90% nucleic acid sequence identity with, even more preferably at least about 95% nucleic acid sequence identity with, and yet more preferably at least about 98% nucleic acid sequence identity with the naturally occurring or wild-type mutagen sensitive gene. Preferably, a variant mutagen sensitive gene hybridizes with the naturally occurring or wild-type mutagen sensitive gene under moderately stringent conditions, more preferably under stringent conditions. Preferably a variant mutagen sensitive gene includes codon and nucleotide substitutions that increase expression or responsiveness in the host cell. Those skilled in the art will appreciate that nucleotide or amino acid changes may alter processing and/or characteristics of the gene.

Variations in the native gene sequence can be made, for example, using various of the techniques known to a skilled worker. Variations can be a substitution, deletion or insertion

of one or more nucleotides. Guidance in determining which nucleotides may be inserted, substituted or deleted without adversely affecting the desired mutagen response may be found by comparing the sequence the gene with that of homologous known genes and minimizing the number of changes made in regions of high homology. The variation allowed may be determined by systematically making insertions, deletions or substitutions of nucleotides in the sequence and testing the resulting variants for the desired mutagen response.

The variations can be made using methods known in the art such as oligonucleotide-mediated mutagenesis, PCR mutagenesis, transposon mutagenesis, mutD mutagenesis, and random chemical or radiation mutagenesis. Site-directed mutagenesis (Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)), cassette mutagenesis (Wells et al., Gene, 34:315 (1985)), restriction selection mutagenesis (Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)) or other known techniques can be performed on the cloned DNA to produce the variant DNA.

Fluorescent Protein

As used herein, "fluorescent protein" refers to any protein capable of fluorescing when excited with appropriate electromagnetic radiation. Fluorescent proteins include proteins having amino acid sequences that are either naturally occurring (wild-type) or variants of these sequences. The fluorescent protein of the invention can fluoresce by any of a variety of mechanisms. Preferably, the fluorescent protein does not require a cofactor or substrate. The fluorescent protein of the invention does not exhibit bioluminescence or chemiluminescence. A preferred fluorescent protein can be employed in a variety of cells.

A fluorescent protein can be isolated from nature or can be produced by recombinant or synthetic means. The term specifically encompasses naturally-occurring truncated or secreted forms of the fluorescent protein, naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of the fluorescent protein.

Fluorescent protein can be measured by a device, such as a fluorimeter (e.g. a fluorometic microtiter plate reader with appropriate filters), a flow cytometer, or by epifluorescence microscopy. Methods for performing assays on fluorescent materials are well known in the art. A change in fluorescence refers to any change in absorption

properties, such as wavelength and intensity, or any change in spectral properties of the emitted light, such as a change of wavelength, fluorescence lifetime, intensity or polarization. Typically, a change in fluorescence refers to a change in output of light.

Green Fluorescent Protein

Suitable fluorescent proteins according to the present invention include green fluorescent protein. As used herein, "green fluorescent protein" (GFP) refers to one of the class of proteins, typically from marine organisms, that emit green light when activated. Suitable green fluorescent proteins have been isolated and characterized from marine organisms such as the Pacific Northwest jellyfish, *Aequorea victoria*, the sea pansy, *Renilla reniformis*, and *Phialidium gregarium*. A preferred green fluorescent protein, such as the GFP of *Aequorea*, can be employed in a variety of cells and requires no substrate to fluoresce. A transformed cell containing the green fluorescent protein fluoresces under various conditions, including when excited with blue light of about 450 nm to about 490 nm. See, e.g., Levine, L. D., et al., Comp. Biochem. Physiol., 72B:77-85 (1982); which is incorporated herein by reference for its disclosure of the structure and manipulation of *Phialidium gregarium* green fluorescent protein.

In a preferred embodiment, the green fluorescent protein is the GFP of the jellyfish *Aequorea victoria* or a variant of this GFP. A variety of useful variants of *Aequorea* green fluorescent protein are known to those of skill in the art. These useful variants have been produced by modifying coding and amino acid sequences of the green fluorescent protein. Residues that can be varied to produce useful variants of *Aequorea* green fluorescent protein include one or more of the amino acids at positions: 26, 64, 65, 66, 68, 72, 73, 99, 100, 123, 127, 145, 146, 147, 148, 149, 153, 154, 163, 164, 167, 185, 202, 203, 208, 212, 235, and 238. Known substitutions at these residues the produce useful variants of *Aequorea* green fluorescent protein include one or more of: K26R, F64L, S65A, S65C, S65G, S65L, S65I, S65T, S65V, Y66H, Y66F, Y66W, V68L, S72A, S73P, F99S, F100S, I123V, K127E, Y145F, Y145H, N146I, N147Y, H148R, N149K, M153T, M154T, V163A, N164H, I167T, I167V, Q185R, S202F, T203I, T203L, T203Y, S208L, N212K, E235D, K238E, and K238N. The variant green fluorescent proteins listed in this paragraph, and polynucleotides encoding

them, can be employed in the methods and compositions of the invention.

Useful variants of green fluorescent protein include those in which at least one of positions 64, 68, or 72 is varied either individually or together with position 65. Useful substitutions at these positions include F64L, V68L, or S72A either individually or together with one of S65T, S65A, or S65G. Useful combinations include F64L and S65T; V68L, S72A, and S65A; and S72A and S65G. By way of further example, useful changes in the excitation and emission spectra of green fluorescent protein can be accomplished through single or combinations of amino acid substitutions including K26R, F64L, S65T, Y66W, N146I, M153T, V163A, N164H, and N212K; F64L and S65T; F64L, Y66H, and V163A; F64L, S65T, Y66H, and Y145F; F64L, S65C, I167T, and K238N; F64L, S65T, F99S, M153T, and V163A; F64L, S65T, F99S, M153T, V163A and S208L; F64L, S65T, Y66W, N146I, M153T, V163A, and N212K; S65A; S65C; S65I; S65L; S65T; S65V; S65G, V68L, S72A, and T203Y; S65T; S65T, M153A, and K238E; S65T, S72A, N149K, M153T, and I167T; Y66H; Y66F; Y66W; Y66H and Y145F; Y66W, M153T, V163A, and N212K; Y66W, N146I, M153T, V163A, and N212K; Y66W, I123V, Y145H, H148R, M153T, V163A, and N212K; S72A, Y145F, and T203L; F99S, M153T, and V163A; F99S, M153T, V163A and S208L; I167V; I167T; and S202F and T203I. It is believed that residues 145-163 are beneficial locations for substitutions that can increase the fluorescence of the *Aequorea* green fluorescent protein. Additional useful changes in the excitation and emission spectra of green fluorescent protein result from combinations of amino acid substitutions such as S73P, F100S, K127E, N147Y, M154T, V164A, and E235D; F100S, M154T, V164A, Q185R, and E235D; and F100S, M154T, and V164A. The variant green fluorescent proteins listed in this paragraph, and polynucleotides encoding them, can be employed in the methods and compositions of the invention.

Amino acid substitutions reported in the art to have no detrimental effect on fluorescence by or other properties of the *Aequorea* green fluorescent protein include changing glutamine-80 to arginine; L3R; D76G; F99I; N105S; E115V; T225S; H231L, and L238E. It is believed that mutations of one or more of residues 76-115 will not adversely affect the function of the *Aequorea* green fluorescent protein. Further, inserting a valine residue as position 2 in a variant extended by one amino acid has no adverse effect on

fluorescence. The variant green fluorescent proteins listed in this paragraph, and polynucleotides encoding them, can be employed in the methods and compositions of the invention.

Single amino acid substitutions that are detrimental to fluorescence intensity when made in the absence of other mutations include S65R, S65N, S65D, S65F, and S65W.

A variety of useful variants of *Aequorea* green fluorescent protein are described in U.S. Patent Nos. 5,625,048, 5,804,287, and 5,998,204, PCT publication WO 96/23810, A. Crameri et al., Nature Biotech. 14, 315-319 (1996), and Cormack et al., Gene 173, 33 (1996); which are incorporated herein by reference for their disclosure of the structure and manipulation of green fluorescent proteins and polynucleotides that encode them. Such variants can be employed in the compositions and methods of the present invention. Methods for manipulating the coding sequence of the *Aequorea* green fluorescent protein, and plasmids including this sequence, are known in the art and are described in, for example, Matthysse, A.G., et al. FEMS Microbiology Letters, 145:87-94 (1996).

A variety of coding sequences for wild type and useful variants of *Aequorea* green fluorescent protein are commercially available, for example, from Clontech (Palo Alto, CA) and Quantum Biotechnologies (Montreal, Quebec, Canada). These suppliers provide the coding sequence in cloning and expression vectors. The coding sequences encode variants of green fluorescent protein that fluoresce blue, cyan, green, and yellow-green, and that can provide brighter fluorescence compared to the wild type protein. Such variants can be employed in the compositions and methods of the present invention.

~~Suitable genes and coding sequences for wild type and variant green fluorescent~~ proteins are described in Prasher et al., Gene 111, 229-233 (1992) and (GenBank Accession No. M62653) and in Figure 4a of U.S. Patent No. 5,958,713. ~~The following~~ sequences can provide a green fluorescent protein coding sequence for ~~the~~ DNA construct of the invention. The nucleotide and corresponding amino acid sequences for an *A. victoria* green fluorescent protein have been reported by Prasher ~~et~~ al. *supra* with GenBank accession numbers M62654 (SEQ ID NO: L) and M62653 and by Inouye et al. (FEBS Lett 351(2-3): 277-280 (1994)) with GenBank accession number L29345. The nucleotide and corresponding amino acid ~~sequences for an~~ *A. victoria* green fluorescent protein mutant 3 ~~have been reported by~~

Cormack et al. (Gene *supra* and Microbiology 143(Part 2):303-11 (1997)) and given GenBank accession number U73901 (SEQ ID NO: J). The disclosures of each of the GenBank accessions mentioned in this paragraph are incorporated herein by reference.

A preferred *Aequorea victoria* green fluorescent protein is the variant encoded by a polynucleotide having the sequence shown in Figure 14 (SEQ ID NO:2), or a degenerate sequence encoding the same amino acid sequence. A preferred degenerate sequence employs codons optimized for expression in the host cell. The amino acid sequence of the preferred green fluorescent protein is also illustrated in Figure 14 (SEQ ID NO: 3). This sequence has been reported as mut2 by Cormack et al. Gene (1996). A portion of this sequence can be PCR amplified using the method of Matthysse et al (1996) using primers such as *gfpHindIII-F* (5'-CTCAAGCTTGATTCTAGATTTAAGAAGG) (SEQ ID No: _____) and *gfpEcoRi-R* (5'-CTCGAATTCTCATTATTTGTATAGTTCATCCATGCC) (SEQ ID No: _____) to generate a 740 base pair product.

Additional Fluorescent Proteins

Other fluorescent proteins can be used in the compositions and methods of the present invention. For example, the fluorescent protein of the invention can be a blue fluorescent protein, which is one of the class of proteins, typically from marine organisms, that emit blue green light when activated. Clontech (Palo Alto, CA) provides cloning and expression vectors encoding a suitable red fluorescent protein found in the IndoPacific sea anemone relative *Discosoma* species. The methods and compositions of the invention can also employ yellow fluorescent protein from *Vibrio fischeri* strain Y-1, phycobiliproteins from marine cyanobacteria such as Synechococcus, e.g., phycoerythrin and phycocyanin, oat phytochromes from oat reconstructed with phycoerythrobilin, or *Propionibacterium freudenreich* uroporphyrinogen III methyltransferase (cobA) gene product. These fluorescent proteins have been described in Baldwin, T. O., et al., Biochemistry 29:5509-5515 (1990); and Wilbanks, S. M., et al., J. Biol. Chem. 268:1226-1235 (1993); which are incorporated herein by reference for their disclosure of the structure and manipulation of fluorescent proteins and polynucleotides that encode them. See also Li et al., Biochemistry 34:7923-7930 (1995) and Wildt S. et al. Nature Biotechnology 17(12):1175-1178 (1999).

Variant Fluorescent Protein

In addition to the full-length native sequence fluorescent protein described herein, it is contemplated that fluorescent protein variants can be employed in the invention. Variant fluorescent proteins remain fluorescent proteins. A "variant" fluorescent protein has at least about 80% amino acid sequence identity with a wild type fluorescent protein. For example, a sequence identity of a variant green fluorescent protein can be compared to a green fluorescent protein having the wild type amino acid sequence, which is known in the art. Such fluorescent protein variants include, for instance, fluorescent proteins wherein one or more amino acid residues are added, or deleted, at the N- or C-terminus, or internally in the sequence. Ordinarily, a fluorescent protein variant will have at least about 80% amino acid sequence identity, preferably at least about 85% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity, even more preferably at least about 95% amino acid sequence identity and yet more preferably 98% amino acid sequence identity with the wild type fluorescent protein.

Fluorescent protein variants can be prepared by introducing appropriate nucleotide changes into the fluorescent protein-encoding DNA, or by synthesis of the desired fluorescent protein. Those skilled in the art will appreciate that amino acid changes may alter processing or characteristics of the fluorescent protein.

Variations in the native full-length sequence fluorescent protein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the fluorescent protein that results in a change in the amino acid sequence of the fluorescent protein as compared with the native sequence fluorescent protein. The variant can also include codons optimized for the host cell. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the fluorescence, or other desired activity, of the fluorescent protein may be found by comparing the sequence of the fluorescent protein with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the

result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for fluorescence, or other desired activity, of the fluorescent protein variant.

The variations can be made using methods known in the art such as oligonucleotide-mediated mutagenesis, PCR mutagenesis, transposon mutagenesis, mutD mutagenesis, and random chemical or radiation mutagenesis. Site-directed mutagenesis (Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)), cassette mutagenesis (Wells et al., Gene, 34:315 (1985)), restriction selection mutagenesis (Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)) or other known techniques can be performed on the cloned DNA to produce the fluorescent protein-encoding variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions (Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)). If alanine substitution does not yield adequate amounts of variant, an isosteric amino acid can be used.

The coding sequence of a wild-type or variant fluorescent protein, such as green fluorescent protein, can be concatenated with those encoding many other proteins; the resulting in a variant fluorescent protein that is a fusion protein. Such fusion proteins typically fluoresce and retain the biochemical features of the partner proteins. See, e.g. Cubitt, A. B., et al., Trends Biochem. Sci. 20:448-455 (1995).

The variant fluorescent protein can also include a localization sequence to direct the indicator to particular cellular sites by fusion to appropriate organellar targeting signals or localized host proteins. A polynucleotide encoding a localization sequence, or signal

sequence, can be ligated or fused at the 5' terminus of a polynucleotide encoding the fluorescence indicator such that the signal peptide is located at the amino terminal end of the resulting fusion polynucleotide or polypeptide. In the case of eukaryotes, the signal peptide is believed to function to transport the fusion polypeptide across the endoplasmic reticulum. The secretory protein is then transported through the Golgi apparatus, into secretory vesicles and into the extracellular space or the external environment. Signal peptides which can be utilized according to the invention include pre-pro peptides which contain a proteolytic enzyme recognition site. Other signal peptides are known to those skilled in the art, or can be readily ascertained without undue experimentation.

The localization sequence can be a nuclear localization sequence, an endoplasmic reticulum localization sequence, a peroxisome localization sequence, a mitochondrial localization sequence, or a localized protein. Localization sequences can be targeting sequences which are described, for example, in "Protein Targeting", chapter 35 of Stryer, L., Biochemistry (4th ed.). W. H. Freeman, 1995. The localization sequence can also be a localized protein.

DNA Construct

The present invention includes a DNA construct with expression of a fluorescent protein operably linked to, an SOS, or other mutagen sensitive, gene. The DNA construct can be a plasmid or other expression vector. Typically, a regulatory region of a mutagen sensitive gene is fused to the coding sequence of a fluorescent protein. Methods for producing, and the structures of, regulatory regions of genes fused to coding sequences are known in the art or can be routinely developed by a skilled worker. For example, a DNA molecule encoding a fluorescent protein, which can be made by PCR, can be ligated to a restriction fragment of a plasmid, vector, or other DNA molecule encoding a mutagen sensitive gene, such as an SOS gene. The coding sequence for the fluorescent protein is positioned so its expression is operatively controlled by the mutagen sensitive gene.

A preferred embodiment of the DNA construct of the invention includes a regulatory region of an SOS gene, such as a promoter of a *umuDC* operon of *E. coli*, fused to the coding sequence of a green fluorescent protein, such as the green fluorescent protein from *Aequorea*.

The this can be accomplished by ligating a PCR fragment encoding a green fluorescent protein to a restriction fragment of a plasmid encoding a umu operon promoter, and all or part of the coding sequences for the umuD gene and/or the umuC gene. A preferred plasmid is plasmid pSE117; Marsh, G.C., et al., J. Bacteriol. 162:155-161 (1985), the disclosure of which is incorporated herein by reference.

~~Figure 15 illustrates another preferred plasmid, pTJgfp. The plasmid includes a~~ preferred umuDC gene (SEQ ID No:1) and a ~~coding sequence~~ for a preferred variant green fluorescent protein (SEQ ID ~~No:3~~). It also includes a colE1 replication origin, ori, and a Bla coding ~~sequence~~ for a β-lactamase selectable marker. The structure and construction of ~~pTJgfp are described in the Examples below and illustrated in Figure 15.~~

The DNA construct typically contains a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*. A preferred selectable marker is a beta-lactamase gene, which provides ampicillin resistance. A variety of selectable markers known to those of skill the art can be routinely employed in the present DNA construct.

The present construct can, alternatively, employ a replicon based on an origin of replication other than ColE1. This is particularly advantageous for making a host cell with more than one construct of the present invention. Two plasmids with the same replication origin cannot be stably maintained in the same cell, since, for example, the enzymatic machinery of the cell does not recognise them as being different and hence all of the copies of both will not necessarily partition evenly into two groups during cell division. Thus, when more than one construct of the invention is employed in a host cell, they must include different origins of replication. Origins of replication different from ColE1 are known in the art and include RP4, R100, pSC101, R6K. Further, a plasmid with an origin of replication different from ColE1 can be advantageous in combination with certain mutagen sensitive DNA polymerase genes. Certain of these genes, such as *polA* genes, are unstable in ColE1 vectors.

For use in mammalian cells, a construct of the invention, or a second construct in the

host cell, can advantageously encode a gene product that can simulate mammalian metabolism, such as a mammalian P450 mixed function oxidase.

Host Cell

A host cell is a cell in which a mutagen sensitive gene can cause production of a fluorescent protein in response to a mutagen. This can be accomplished by a variety of different cells, for example prokaryotic cells and eukaryotic (e.g. mammalian) cells, or cells of single or multi- cellular organisms. Preferably, the host cell is a prokaryotic cell, preferably a bacterial cell. The host cell is suitable for expression and induction of the DNA construct. Suitable host cells include Enterobacteriaceae (e.g. *Shigella, Salmonella, Escherichia, Proteus, Erwinia* and the like) and gram positive bacteria (e.g. *Bacillus, Streptococcus, Staphylococcus*, the *Actinomycetes*, and the like). Preferred host cells include a cell in which an SOS, or other mutagen sensitive, gene can be expressed when the host cell's DNA is damaged by a mutagen. A preferred host cell is a microbe, preferably a bacterium, subject to or capable of an SOS response.

Preferred bacteria host cells include *Escherichia coli*, such as an *E.coli* K12, B or B/r strain; and including *E. coli* strains AB1157, DB6659, GW2100, DH5, CC118, CSH26; and the like. Preferred bacteria host cells include *Salmonella typhimurium*, such as the *S. typhimurium* LT2 subline strains developed in the Ames lab at Berkeley, e.g. TA1535, TA1537, TA1537, TA1538, TA98, TA100, TA102 and subsequent variations or improvements of them; a *S. typhimurium* strain from the LT2 or LT22 sublines; a *S. typhimurium* strain NM1011, NM2009, NM5004, NM6001, NM6002, or OY1001/1A2; and the like. Certain strains of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Bacillus subtilis*, *Neurospora crassa*, or *Pseudomonas aeuroginosa* can also be employed as host cells to carry a variety of the plasmid variants as previously described.

Suitable host cell strains are described in Walker (1984), *supra*; Marsh, G.C., et al. (1985) *supra*; Herrerro, M., et al. J. Bacteriol. 172:6557-6567 (1990); Justus, T. and Thomas, S.M. Mutation Res. 398: 131-141 (1998); and Justus, T. and Thomas, S.M. Mutagenesis 14(4):351-356 (1999); which are incorporated herein by reference for their disclosure of characteristics and manipulation of these host cell strains. See also, Miller, J. H.,

Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, 1972; Oda Y. et al. Mutation Research 334(2):145-56 (1995); Oda Y. et al. Mutation Research 272(2):91-9 (1992); Aryal P. et al. Mutation Research-Genetic Toxicology & Environmental Mutagenesis 442(2):113-120 (1999); Oda Y. et al. Carcinogenesis 20(6):1079-1083 (1999); Oda Y. et al. Carcinogenesis 17(2):297-302 (1996); and Schmid C. et al. Mutation Research-Genetic Toxicology & Environmental Mutagenesis 394(1-3):9-16 (1997).

A host cell may be cultured under standard conditions of temperature, incubation time, optical density, plating density and media composition corresponding to the nutritional and physiological requirements of the host cell. However, conditions for maintenance and growth of the host cell may be different from those for assaying candidate mutagens in the methods of the invention. Modified culture conditions and media can be used to facilitate detection of the expression of a fluorescent protein or an SOS response.

Host cell strains, cultures, or cell lines may be expanded, stored, or retrieved by a variety of techniques known in the art and appropriate to the host cell. For example, a host cell of the invention can be preserved by stab culture, plate culture, or in glycerol suspensions and cryopreserved in a freezer (at -20.degree. C. to -100.degree. C.) or under liquid nitrogen (-176.degree. C. to -196.degree. C.).

Selection and Transformation of Host Cells

Host cells are transfected or transformed with a DNA construct described herein for mutagen detection, an SOS response, or fluorescent protein expression and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH, and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Hanahan, D. (1985) Techniques for transformation of *E.coli*; In Glover, D.M. (ed) "DNA Cloning Vol1: a practical approach", IRL Press, Oxford, UK, pp109-135. A standard method for the transformation of plasmids into *E.coli* uses cold calcium chloride. Greater competence can also be generated through the use of more involved protocols known in the art and employing rubidium

chloride or hexamine cobalt chloride.

Additional methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ treatment and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook *et al.*, *supra*, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers.

Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw *et al.*, Gene, 23: 315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen *et al.*, J. Bact., 130: 946 (1977) and Hsiao *et al.*, Proc. Natl. Acad. Sci. (USA), 76: 3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene or polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown *et al.*, Methods in Enzymology, 185: 527-537 (1990) and Mansour *et al.*, Nature, 336: 348-352 (1988).

Detecting Gene Amplification/Expression

A host cell which contains a sequence encoding a fluorescent protein or an SOS gene and which express the fluorescent protein or the SOS gene may be identified by a variety of known methods. For example the host cell can be assayed for the presence or absence of "marker" gene functions (e.g., resistance to antibiotics) or detection of fluorescent protein by, for example, its fluorescence (e.g. detection of fluorescence by epifluorescence microscopy). Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA (Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)), dot blotting (DNA analysis), *in situ* hybridization, using an appropriately labeled probe, or PCR based on appropriate

primers, based on the sequences described herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal.

Detecting Mutagens

A mutagen can be detected by incubating a host cell in a suitable nutrient medium, such as nutrient broth or on nutrient agar. The nutrient medium can include a suitable selection reagent, such as an antibiotic, to select for host cells including a DNA construct of the invention. Stationary phase host cells can be obtained by growing the cells for a prolonged period, such as overnight. Logarithmic phase host cells can be obtained by diluting stationary cultures and incubating them again before use for detecting a mutagen. For example, logarithmic phase cells can be obtained by diluting overnight cultures cells 1:10 and regrowing them for two hours at 37 °C before use.

For testing a chemical mutagen, a test compound is added to a host cell culture and the cell culture is incubated for a time suitable for appearance of a fluorescent signal in the presence of a mutagen. In a preferred embodiment, the amount of a test compound added to a cell culture does not adversely effect culture conditions for growth of the host cell. For example, about 0.01 to about 0.1 ml of a test compound solution can be added to about 2 ml of a host cell culture. Alternatively, in an embodiment preferred for detecting weak mutagens or mutagens in an environmental sample, sufficient nutrient media can be displaced or replaced by test compound that the host cell is deprived of nutrient to some degree, or even starved.

During treatment with mutagen, the cells can be maintained either at stationary or logarithmic phase. Preferably, if cells are deprived of nutrients or starved, they are maintained at logarithmic phase during exposure to mutagen. A host cell can be incubated with a mutagen for a time as short as 1 or 2 hours or as long as about 24 or 48 hours, or longer. Preferred incubation times include about 4.5 to about 24 hours. Advantageously, stability of the fluorescence produced by a fluorescent protein, such as green fluorescent protein, allows the present method to be conducted with incubations longer than can be achieved with previous systems that detect mutagens by producing light or a visible signal.

After treatment with mutagen, the cells can be diluted with nutrient medium and incubated for the detection of mutagenesis. For example, a 2 ml mutagen treated host cell sample can be diluted with 18 ml of nutrient medium. The host diluted cells are incubated under conditions suitable for expression of the fluorescent protein. For example, the diluted host cells can be dispensed at 0.15 ml/well into 96 well microtiter plates and incubated at 37 °C. Alternatively, a sufficient number of cells to start a new culture can be removed from the mutagen treatment mixture and used to inoculate an aliquot of nutrient medium. For example, about 10 to about 50 cells can be employed to inoculate a 0.15 ml culture in a microtiter plate. The plate and cells can then be incubated at 37 °C.

After the host cells have been grown under conditions suitable for expression of the fluorescent protein, fluorescence can be monitored using any suitable apparatus, several of which are listed above. For example, fluorescence can be monitored by employing a fluorescence microtiter plate reader with the mission and excitation wavelengths selected for the fluorescent protein. Green fluorescent protein can be detected with an emission wavelength of about 510 or about 520 nm and an excitation wavelength of about 485 nm. Appropriate excitation and emission wavelengths for existing variant green fluorescent proteins are known in the art. An increase in fluorescence output, level, or distribution compared to a suitable control indicates that a test compound is a mutagen. Fluorescence can be read directly and/or automatically from the culture of host cells, which provides an advantage in steps, time, materials, and cell handling compared to previous methods that require culturing revertant colonies, and like methods.

A mutagen typically produces a dose dependent increase in fluorescence, although

low levels of some mutagens may not produce an increase that is apparent without statistical analysis. Similarly, some mutagens may require a longer or shorter time than others to produce a significant or dose dependent increase.

Suitable nutrient growth media for bacteria that can be employed during incubation to detect mutagens include nutrient broth, Luria broth, defined minimal media, such as Vogel-Bonner or Davis-Mingoli salts.

Detecting Antimutagens

Antimutagens are agents that lower a mutation rate. They have potential as therapeutic agents to control proliferative conditions in individuals affected by DNA repair deficiencies. Their detection is problematic in conventional assays that rely on the detection of a lowering of mutant colony counts in assays such as the Ames test. The present construct and host cell can test for antimutagens that, when added to cultures of host cells containing the present construct, lower fluorescent output. For example, the cells can be exposed to both a mutagen and an antimutagen simultaneously, and the fluorescent output is lower than from cells exposed only to the mutagen. A concurrent toxicity assay can check that the decrease in fluorescence was not due to cell death.

Toxicity Assays

When testing mutagens or antimutagens, a toxicity assay can be run concurrently, on the same or different cells. A toxicity assay can provide a suitable control to account for any decrease in fluorescence that might be due to general toxicity to the host cells.

Statistical Analysis

The present DNA construct, host cell, and methods can be employed to a detecting mutagen solely on the basis of an increase in the level of fluorescent protein expression or detected fluorescence compared to a suitable control. Statistical analysis of the location, distribution, amount, pattern, or the like of fluorescent protein can be conducted on replicate mutant treated samples. For example, under conditions where there is no significant difference between the arithmetic mean of samples, statistical analysis of the distribution of

fluorescence can detect a mutagen.

For accomplishing such a statistical analysis in standard assays such as the Ames test, multiple replicate cultures (either in the presence or absence of a mutagen) can be grown and a sample of each analysed to see how many mutants it contains (as measured by a plating assay for mutant colonies). The resulting mathematical distribution describes the numbers of mutants in many cultures. Changes in the shape of the distribution or its placement measure far subtler changes in mutation rate than a simple shift in the arithmetic mean. However, analyses of this type are employed infrequently due to their technical difficulty

A fluorescent protein, such as green fluorescent protein can be employed in an assay of this type to determine changes in the shape or placement of a mutant distribution from replicate cultures without increasing the technical difficulty of the experiment. Interestingly, the fluorescence output of any individual culture carrying a mutagen sensitive gene operably linked to a fluorescent protein may not correlate with the number of mutants in that culture, since transcription of a single fused gene need not mirror the whole DNA repair capacity of a growing culture. However, a stable fluorescent signal, such as that from green fluorescent protein provides a cumulative record of transcription of a single fused gene, and indicates how this varies between replicate cultures. Significantly, this type of analysis does not work with the *lux* reporter. Thus, the present construct can be used to detect subtle mutagens that otherwise might be missed in conventional assays, as well as more potent ones.

In one embodiment, the Kolmogorov-Smirnov Z Test (KZ test) can be employed to detect differences in both the locations and shapes of fluorescence distributions. This test is well characterized for analysis of variable locations, shapes, and distributions of data. A significant difference between a control and a test compound according to the Kolmogorov-Smirnov Z test indicates that the test compound is a mutagen. A significant difference can be denoted by a P value of less than about 0.1, preferably less than about 0.05. The Kolgomogorov-Smirnov Z test can be applied according to the SPSS statistical analysis software package, version 7.5. The SPSS software package also includes the Mann-Whitney test, which can illustrate the differences in distribution.

The present invention may be better understood with reference to the following

examples. These examples are intended to be representative of specific embodiments of the invention, and are not intended as limiting the scope of the invention.

Examples

Example 1 - - Transcriptional Fusions of Green Fluorescent Protein as Reporters of Bacterial Mutagenesis

A fluorescent protein under control of a mutagen sensitive gene produces a robust and dose-dependent response to mutagens.

Materials and Methods

Bacterial strains and plasmids. Bacterial strains and plasmids used in these examples are listed in Table 1.

Table 1. Bacterial strains used in this study

Strain	Relevant characteristics	Source/reference
E. coli GW2100	AB1157 *umuC*::Tn5	G. Walker
E. coli DH5/pSE177	*recA* $hsdR_{EcoK}$/pSE117	Marsh and Walker (1985)
E. coli GW2100/pTJ10	AB1157 *umuC*::Tn5/pTJ10 (*umuC'*::*luxAB*)	Justus and Thomas (1998)
E. coli GW2100/pTJ*gfp*	AB1157 *umuC*::Tn5/pTJ*gfp* (*umuC'*::*gfp*)	This example
S. typhimurium TA1537	*HisC*3076Δ (*gal uvrB bio*) *rfa*	B.N. Ames via D.G. MacPhee
S. typhimurium TA1537/pTJ10	*HisC*3076Δ (*gal uvrB bio*) *rfa*/TJ10 (*umuC'*::*luxAB*)	This example

Media and growth conditions. Bacteria were grown in nutrient broth (Oxoid CM1) or on nutrient agar (Oxoid CM3), both supplemented with the addition of 50 μg/ml ampicillin (Sigma A951), and shaken at 200 r.p.m. at 27°C.

DNA cloning. Methods for plasmid DNA isolation, restriction endonuclease digests, DNA ligation, alkaline phosphatase treatment and agarose gel electrophoresis techniques followed standard protocols. Cells were transformed by standard methods with the following modifications: cells were heat shocked at 37°C for 10 min and then grown in nutrient broth. DNA was isolated from agarose using agarase from *Pseudomonas atlantica* according to the

manufacturer's instructions (Boehringer Mannheim).

PCR. PCR amplification of the *gfp* gene sequence was conducted as described by Matthysse et al., FEMS Microbiol. Lett. 145:87-94 (1996), the disclosure of which is incorporated herein by reference for disclosure of PCR and microscopy methods.

umuC-gfp gene fusion construction. The 700 bp *gfp* PCR fragment was ligated to a 7 kb *Hind*III-*Eco*RI fragment from plasmid pSE117 (Figure 1). *Escherichia coli* GW2100 cells were transformed with this ligation mixture and transformants selected on nutrient agar with ampicillin.

gfp-umuC gene fusion confirmation. Plasmid preparations of colonies that resulted from ligation were PCR amplified using *gfp* primers to confirm the presence of the *gfp* gene (data not shown). Colonies were also screened for fluorescence using epifluorescence microscopy.

Fluorescent detection methods. Stationary phase cells were obtained for experiments by allowing cells to grow up overnight. Logarithmic phase cells were obtained by diluting overnight culture cells 1:10 and regrowing them for 2 h at 37°C prior to use in experiments. During fluorescent assays bacterial cell culture in either logarithmic phase or stationary phase (2 ml) were with incubated for a predetermined time with mutagen, and then added to 18 ml of nutrient broth or 0.85% saline with ampicillin. For UV-irradiated cells this involved irradiating cells in Falcon 60x15 mm tissue culture dishes with regular stirring. All other mutagens were added to this 2 ml of cells. This mixture was then dispensed at 150 µl/well into 96 Uniwell microtiter plates (Elkay) and incubated at 37° C with shaking. Fluorescence was measured using a Fluoroskan Ascent by Labsystems using the Ascent computer software package v.2.2, with an emission filter of 510 nm and excitation filter of 485 nm. Measurements were taken for three incubation times with mutagen; time 1 being fluorescence prior to incubation, time 2 at 4.5 h, and time 3 at 24 h. All fluorescent values are presented as the means (error bars show standard deviations) of 96 cultures grown in parallel to obtain a statistically more accurate representation of mutation rate.

Results

Cultures in both logarithmic and stationary growth phases were tested following

mutagen treatment (Figures 2-4). The fluorescent signal from stationary phase cells did not differ significantly from those in logarithmic phase (Figures 2-4). The chemical mutagen *N*-methyl-*N*-nitro-*N*-nitrosoguanidine (MNNG) induced a dose-dependent increase in fluorescence levels at 1.5, 3.5 and 7 μg/ml (Figure 2). In this study, there was no significant increase in fluorescence level at 0.1 μg/ml when compared with the 0 dose control (Figure 2). When another chemical mutagen, methylmethane sulphonate (MMS), was assayed cells showed no significant difference in fluorescence level from the 0 dose control and 13 μg/ml MMS, but they responded in a dose-dependent manner at 325 and 650 μg/ml (Figure 3). At a higher dose of 1300 μg/ml a lower level of fluorescence was observed, most likely due to cellular toxicity (Figure 3). The physical mutagen UV irradiation failed to induce a significant difference in fluorescence level from the 0 dose control at 1 $J/M^{2,}$ but at 3 and 6 J/M^2 a dose-dependent increase in fluorescence level was seen (Figure 4).

When assays were conducted over periods of time up to 24 h the signal from the *gfp* reporter system increased with time whilst there was a marked decrease in *lux*-derived luminescence (Figures 2-4). This demonstrates the more robust nature of the *gfp* system compared with *lux*, which is very dependent on cell physiology at the immediate time of assay due to its shorter half-life. *gfp* is thus more suitable than the luciferase system for a variety of assays, including those where the mutagen dose being assayed is comparatively high and/or a rapid result is not crucial.

gfp gene expression cells were assayed in 0.85% saline. The results shown in Figures 5 and 6 indicate that cells do not require cultivation in a nutrient-rich medium for a dose-dependent response to a mutagen to occur. However, there is the requirement that cells be in logarithmic phase (Figure 5) and not a stationary phase (Figure 6) for reporter gene induction to occur in saline.

Discussion

Previous work has shown that the *luxA* and *luxB* (luciferase) genes from *Vibrio harveyi* can be successfully fused to bacterial DNA repair genes to provide a sensitive and semi-quantitative assay for the presence of a mutagen. One drawback with this luciferase

system, however, is the need for cells to be in a metabolically active state, together with the transient state of the luminescent signal that begins to decline after 24 h. This means that for the test to be useful samples would need to be processed in the laboratory under quite invariant conditions.

A distinct advantage of the *gfp* system though does appear to be the robust and long lasting nature of the signal. When cells carrying either reporter construct were left in the presence of mutagens for 24 h before reading, fluorescent signals could still be detected easily, whereas luminescence was not. Thus *gfp*-based testing can be employed under conditions where extended exposure to mutagenic agents is required.

The *gfp*-based reporter system showed no detectable variation in response when using cells from different growth phases in nutrient broth, unlike the *lux* system. Dose-dependent increases in fluorescence were observed for each of the mutagens tested. However, when cells were starved, induced fluorescence was only seen in logarithmic phase cells. This is another advantage of the *gfp* reporter system compared with the *lux*. As a result, green fluorescent protein may be a more suitable reporter for detection of high levels of mutagens in environmental samples, as often nutrient levels in such samples may be low.

Example 2 - - Transcriptional Fusions of Green Fluorescent Protein for Detecting Weak Mutagens

A fluorescent protein under control of a mutagen sensitive gene produces a robust and dose-dependent response to weak mutagens and low concentrations of mutagens.

Materials and Methods

Bacterial strains and plasmids. The bacterial strain used in this study was *E. coli* GW2100/pTJ*gfp*, the details of which are described above in Example 1. Plasmid pTJ*gfp* contains a transcriptional fusion between the promoter of the *umuC* DNA repair gene of *E. coli* , and the *gfp* fluorescence gene of *Aqueorea victoria*. It has been shown in Example 1 to confer on cells the ability to emit light when challenged with DNA damaging agents. GW2100 is a derivative of the standard AB1157 strain in which transposon Tn5 insertion has inactivated the chromosomal *umuC* gene.

Detection of mutation from *arg-3* to Arg^+ using colony assays. Cultures of *E. coli* were grown overnight in Nutrient Broth (Oxoid Code CM1) at 37°C, shaking at 200rpm. The resultant stationary phase cultures were diluted by 10^8 and, where appropriate, mutagens were added. 150µl aliquots of the diluted cultures were then added to each of 24 or 48 microtiter plate wells and incubated at 37° C overnight. Following incubation a 100µl sample was removed from each well and plated onto Davis Mingioli Minimal O Medium, supplemented with 50µl histidine, proline, leucine and threonine, 5µl/ml thiamine, and with glucose (1%w/v) present as a sole carbon source. Plates were incubated for four days and then scored for Arg^+ mutant colonies. Viable counts were obtained by plating 100µl of suitably diluted samples from four randomly chosen wells onto nutrient agar, followed by overnight incubation at 37°C.

Detection of fluorescence emission from fusion strain cultures in microtiter plates. Culture growth conditions and experimental protocols were performed generally as described above in Example 1. Between 10 and 50 cells from a suitably diluted logarithmic phase culture were used to inoculate parallel 150µl cultures in 96 well flat bottomed microtiter plates (Elkay), and incubated at 37° C with shaking. Fluorescence was quantified using a Fluoroskan Ascent by Labsystems, using an emission filter of 520nm and an excitation filter of 485nm.

Statistical analysis. Statistical analysis of multiple parallel cultures to detect and compare low levels of mutation were performed using the SPSS statistical analysis software package, version 7.5. The Kolmogorov-Smirnov Z Test (KS Test), which detects differences in both the locations and shapes of data distributions, was used to analyze the fluorescent values obtained from microtiter plate assays after 24 hours incubation and the mutant colony distributions obtained after 4 day plate incubations. In all cases the zero dose control distribution was compared to distributions obtained in the presence of exogenous mutagen.

Results

Detection of chemical mutagens by revertant colony assay.

Reversion of the *arg-3* marker in *E. coli* strain GW2100/pTJ*gfp* was analyzed in

approximately 40 parallel cultures exposed to varying doses of MNNG (Figure 7) and MMS (Figure 8). Exposure to 3.5μg/ml MNNG elicited a positive response as seen by a significant increase in the mean revertant colony numbers of the 40 exposed cultures compared to the 40 zero exposure controls (Figure 7). Likewise exposure of parallel cultures to 325μg/ml MMS gives a clearly positive result where the mean number of revertant colonies per culture sample changes, and while this change does not occur at 13μg/ml, mutagenicity is still detectable through a change in the culture to culture variation in mutant numbers (Figure 8).

Direct comparison of revertant colony assay and fluorescent emission from a gene fusion.

In a single experiment, 24 parallel cultures of GW2100/pTJ*gfp* were exposed to 1.7 μl of MNNG and both mutation at the *arg-3* chromosomal locus versus transcription of the plasmid-encoded *umuC'-gfp* fusion analyzed (Figures 9a and 9b). Note from the data that the mutagenic activity of the chosen dose of MNNG can be detected by a shift in the culture to culture variation and arithmetic mean of either revertant colony numbers (Figure 9b) or fluorescent emission (Figure 9a) compared to unexposed controls. The latter experiment represents a substantial technical simplification to measurement protocols provided through the inherent stability of the *gfp* gene product. Instead of requiring the significant agar medium preparation and sample processing of a plate count, the experiment was run in a single 96 well microtiter plate and read automatically.

Confirmation of the effectiveness of the gene fusion strain in detecting other mutagens.

The ability of the *umuC'-gfp* fusion plasmid to detect mutagenic activity of other potent agents was confirmed. In Figure 10 a total of 84 parallel cultures were exposed to each of 0, 13 and 325μg/ml of the alkylating agent methylmethane sulphonate (MMS). Note that at 325μg/ml mutagenic activity of the compound is pronounced, as shown by a substantial shift in the mean fluorescent output per culture compared to the unexposed control series. Visually there is a minor shift in the distribution of fluorescent output from cultures exposed to 13μg/ml compared to the control series, although the mean fluorescent output of this bank of cultures is similar to the unexposed cultures. This difference was

analyzed statistically as detailed below.

In a similar vein a parallel series of 84 replicate cultures were exposed to each of 0, 0.1 and 3.5μg/ml MNNG (Figure 11). Mutagenic activity was clearly evident at 3.5μg/ml as shown by the significant shift in the distribution of fluorescent output per culture compared to the unexposed controls. At 0.1 μg/ml a minor shift in culture to culture variation of fluorescent output was evident, and was again analyzed statistically. Finally 84 parallel cultures were exposed to 0, 1 and 3 J/m^2 of 254nm ultraviolet light (UVC) (Figure 12). The mutagenic activity of UVC is clearly evident at 3 J/m^2, and the minor shift evident in cultures exposed to 1 J/m^2 was analyzed statistically. Thus the quantification of fluorescent output can be successfully used to assess mutagenicity of both physical and chemical agents.

Statistical comparison of fluorescent emission patterns from different treatments.

Table 2 shows results of use of the KS test in detecting differences in mutation rate between unexposed banks of parallel cultures versus those exposed to low level mutagens (13 μg/ml MMS, Figure 10; 0.1μg/ml MNNG, Figure 11; and 1 J/m^2, Figure 12). P values of less than 0.05 result in rejection of the null hypothesis that the two fluorescence distributions being compared are similar, thus indicating a significant difference in the pattern and or location of the fluorescence output from each treatment. Although there was no marked difference between the arithmetic mean of the fluorescent output of 84 replicate cultures exposed to low doses of MMS (13.22 at 13μg/ml, Figure 10), MNNG (10.82 at 0.1μg/ml, Figure 11) or UV (14.37 at 1J/m^2, Figure 12) compared to similar numbers of unexposed cultures (11.99 at 0 dose treatments in Figures 10-12), statistical analysis using the KS test confirmed that the distributions were significantly different from the control. Similar analyses confirmed the existence of such differences in banks of cultures exposed to higher mutagen doses that are detectable by standard colony reversion assays (Table 3). Thus, fluorescent emission from a transcriptional fusion strain can also be used to detect and quantify the effects of low dose (or weak) mutagens.

Table 2 Summary statistics used in comparing responses to MMS, MNNG and UV exposure, as analysed by fluorescence emission

Dose	No mutagen	UV ($J.M^2$)		MMS (μg/ml)		MNNG (μg/ml)	
		1	3	13	325	0.1	3.5
Mean	11.99	14.37	17.41	13.22	36.5	10.82	19.39
S.D	0.91	0.78	1.42	1.28	3.42	0.61	1.53
N	84	84	84	84	84	84	84
K.S statistic [A]		5.4	6.33	3.09	6.48	3.94	6.48
Asymp. Sig. (2 tailed) [B]		<0.000	<0.000	<0.000	<0.000	<0.000	<0.000

A - test of each mutagen versus the control; B - P values derived from the Kolmogorov-Smirnov 2 sample test

Table 3 Summary statistics used in comparing responses to MMS and MNNG exposure, as analysed by revertant colony numbers

Dose	No Mutagen	MMS (μg/ml)		No Mutagen	MNNG (μg/ml)	
		13	325		0.1	3.5
Mean	3.53	8.84	17.95	4.1	8.05	24.9
S.D	0	5.68	11.22	4.96	7.7	38
N	37	37	37	40	40	40
K.S statistic [A]		2.711	3.889		1.789	3.130
Asymp. Sig. (2 tailed) [B]		<0.000	<0.000		<0.003	<0.000

A - test of each mutagen versus the control; B - P values derived from the Kolmogorov-Smirnov 2 sample test

Discussion

Traditionally revertant colony counts are widely used to detect mutagenicity and provide a satisfactory tool for the analysis of potent agents. This study has shown that a transcriptional fusion strain in which an inducible DNA repair gene involved in the mutagenesis process has been fused to a reporter for fluorescent emission can also be used to detect agents that are mutagenic. Validation of the strain and the method have been presented using the known mutagens MMS, MNNG and UVC. The strain is simple to use, requires less preparation or technical proficiency during experimental execution, and yields sensitive results. The stable nature of the fluorescent protein product of the *Aequerorea*

victoria gfp gene allows the researcher to examine the cumulative history of DNA repair gene transcription in a culture in the same way that a revertant colony count provides an end-point analysis of mutational events. The difference between the two systems may be that increased transcription of DNA repair genes is observed at lower doses than those required to induce reversion in a single gene marker. The results of this study show there is a lowering in the threshold dose of potentially hazardous chemicals that can be adequately detected by the fluorescence emission method compared to using duplicate or triplicate plate counts. The improved method thus also provides a capacity to detect weaker agents than have been analyzed to date; weaker mutagens could be less toxic and therefore escape detection in other cellular assays.

Other reporter genes whose products are less stable than the *gfp* used here (e.g. the *lux* genes encoding bacterial luciferase) are less amenable to this type of analysis (data not shown) because they respond to more general physiological changes such as temperature, pH etc. that may confound experimental results.

It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

All publications and patent applications in this specification are indicative of the level of ordinary skill in the art to which this invention pertains.

The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention.